

# EDS ASSIGNMENT

Name :- Vaishnavi Aher

Class :- CS8

Roll No. :- 61

PRN :- 202401120002

---

Dataset Name :- Paper Review

# Problem Statement

```
eds.py > ...
1 import pandas as pd
2 import numpy as np
3
4 # Step 1: Create the dataset
5 data = {
6     'Paper_ID': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
7     'Reviewer_ID': [101, 102, 103, 101, 104, 102, 105, 106, 101, 107],
8     'Review_Score': [4.2, 3.5, 4.8, 4.0, 4.5, 3.0, 4.7, 4.3, 3.8, 4.6],
9     'Review_Date': ['2025-04-25', '2025-04-24', '2025-04-23', '2025-04-22', '2025-04-21', '2025-04-20', '2025-04-19', '2025-04-18', '2025-04-17', '2025-04-16'],
10    'Paper_Title': ['AI in Healthcare', 'Quantum Computing Applications', 'Data Science in Business', 'Blockchain for Security', 'Robotics in Manufacturing',
11                  'AI in Healthcare', 'Machine Learning for Cybersecurity', 'Cloud Computing in Education', 'Data Science in Business', 'Blockchain for Security'],
12    'Reviewer_Affiliation': ['University A', 'University B', 'University C', 'University A', 'University D', 'University B', 'University E', 'University F', 'University A', 'University'],
13    'Reviewer_Experience (years)': [10, 15, 5, 10, 12, 15, 8, 6, 10, 18],
14    'Paper_Category': ['AI', 'Computer Science', 'Data Science', 'Blockchain', 'Robotics', 'AI', 'Machine Learning', 'Cloud Computing', 'Data Science', 'Blockchain']
15 }
16
17 # Load dataset into DataFrame
18 df = pd.DataFrame(data)
19
20 # Convert Review_Date to datetime for easier manipulation
21 df['Review_Date'] = pd.to_datetime(df['Review_Date'])
22
23
24
25
26 # Step 2: Answering the 20 problem statements
27
28 # 1. Find the average review score for each paper.
29 avg_score_per_paper = df.groupby('Paper_ID')['Review_Score'].mean()
30 print("1. Average review score for each paper:")
31 print(avg_score_per_paper)
```

## Output

```
PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL PORTS
● PS C:\Users\De11\Desktop\EDS> python eds.py
1. Average review score for each paper:
Paper_ID
1      4.2
2      3.5
3      4.8
4      4.0
5      4.5
6      3.0
7      4.7
8      4.3
9      3.8
10     4.6
Name: Review_Score, dtype: float64
```

```

33 # 2. Find the highest review score for each reviewer.
34 highest_score_per_reviewer = df.groupby('Reviewer_ID')['Review_Score'].max()
35 print("\n2. Highest review score for each reviewer:")
36 print(highest_score_per_reviewer)
37
38 # 3. Count the number of papers reviewed by each reviewer.
39 papers_reviewed_by_each = df.groupby('Reviewer_ID')['Paper_ID'].count()
40 print("\n3. Number of papers reviewed by each reviewer:")
41 print(papers_reviewed_by_each)
42
43 # 4. Find the paper with the highest review score.
44 highest_reviewed_paper = df.loc[df['Review_Score'].idxmax()]
45 print("\n4. Paper with the highest review score:")
46 print(highest_reviewed_paper)
47
48 # 5. Find the paper with the lowest review score.
49 lowest_reviewed_paper = df.loc[df['Review_Score'].idxmin()]
50 print("\n5. Paper with the lowest review score:")
51 print(lowest_reviewed_paper)
52
53 # 6. Calculate the average experience of all reviewers.
54 avg_experience = df['Reviewer_Experience (years)'].mean()
55 print("\n6. Average experience of all reviewers:")
56 print(avg_experience)
57

```

## Output

```

2. Highest review score for each reviewer:
Reviewer_ID
101    4.2
102    3.5
103    4.8
104    4.5
105    4.7
106    4.3
107    4.6
Name: Review_Score, dtype: float64

3. Number of papers reviewed by each reviewer:
Reviewer_ID
101     3
102     2
103     1
104     1
105     1
106     1
107     1
Name: Paper_ID, dtype: int64

4. Paper with the highest review score:
Paper_ID          3
Reviewer_ID       103
Review_Score      4.8
Review_Date       2025-04-23 00:00:00
Paper_Title       Data Science in Business
Reviewer_Affiliation  University C
Reviewer_Experience (years)  5
Paper_Category    Data Science
Name: 2, dtype: object

5. Paper with the lowest review score:
Paper_ID          6
Reviewer_ID       102
Review_Score      3.0
Review_Date       2025-04-20 00:00:00
Paper_Title       AI in Healthcare
Reviewer_Affiliation  University B
Reviewer_Experience (years)  15
Paper_Category    AI
Name: 5, dtype: object

6. Average experience of all reviewers:
10.9

```

```

Click to add a breakpoint
58 # 7. Calculate the average review score per paper category.
59 avg_score_per_category = df.groupby('Paper_Category')['Review_Score'].mean()
60 print("\n7. Average review score per paper category:")
61 print(avg_score_per_category)
62
63 # 8. List the papers reviewed by a specific reviewer (e.g., Reviewer 101).
64 papers_by_reviewer_101 = df[df['Reviewer_ID'] == 101]
65 print("\n8. Papers reviewed by Reviewer 101:")
66 print(papers_by_reviewer_101)
67
68 # 9. Find the reviewer with the highest average review score.
69 highest_avg_score_reviewer = df.groupby('Reviewer_ID')['Review_Score'].mean().idxmax()
70 print("\n9. Reviewer with the highest average review score:")
71 print(highest_avg_score_reviewer)
72
73 # 10. Find the total number of papers for each paper category.
74 papers_count_per_category = df.groupby('Paper_Category')['Paper_ID'].count()
75 print("\n10. Total number of papers in each category:")
76 print(papers_count_per_category)
77
78 # 11. Find the correlation between the review score and reviewer experience.
79 score_experience_corr = df[['Review_Score', 'Reviewer_Experience (years)']].corr()
80 print("\n11. Correlation between review score and reviewer experience:")
81 print(score_experience_corr)
82
83 # 12. Find the papers that were reviewed on or after 2025-04-20.
84 papers_reviewed_after_2025_04_20 = df[df['Review_Date'] >= '2025-04-20']
85 print("\n12. Papers reviewed on or after 2025-04-20:")
86 print(papers_reviewed_after_2025_04_20)
87
88 # 13. Count the number of papers in each category (e.g., AI, Data Science).
89 paper_count_category = df['Paper_Category'].value_counts()
90 print("\n13. Count of papers in each category:")
91 print(paper_count_category)
92

```

# Output

```

7. Average review score per paper category:
Paper_Category
AI          3.6
Blockchain  4.3
Cloud Computing  4.3
Computer Science  3.5
Data Science  4.3
Machine Learning  4.7
Robotics     4.5
Name: Review_Score, dtype: float64

8. Papers reviewed by Reviewer 101:
   Paper_ID  Reviewer_ID  Review_Score  Review_Date  Paper_Title  Reviewer_Affiliation  Reviewer_Experience (years)  Paper_Category
0         1         101         4.2   2025-04-25      AI in Healthcare      University A              10             AI
3         4         101         4.0   2025-04-22  Blockchain for Security      University A              10          Blockchain
8         9         101         3.8   2025-04-17  Data Science in Business      University A              10          Data Science

9. Reviewer with the highest average review score:
103

10. Total number of papers in each category:
Paper_Category
AI          2
Blockchain  2
Cloud Computing  1
Computer Science  1
Data Science  2
Machine Learning  1
Robotics     1
Name: Paper_ID, dtype: int64

11. Correlation between review score and reviewer experience:
              Review_Score  Reviewer_Experience (years)
Review_Score      1.000000              -0.446163
Reviewer_Experience (years)  -0.446163              1.000000

12. Papers reviewed on or after 2025-04-20:
   Paper_ID  Reviewer_ID  Review_Score  Review_Date  Paper_Title  Reviewer_Affiliation  Reviewer_Experience (years)  Paper_Category
0         1         101         4.2   2025-04-25      AI in Healthcare      University A              10             AI
1         2         102         3.5   2025-04-24  Quantum Computing Applications      University B              15          Computer Science
2         3         103         4.8   2025-04-23      Data Science in Business      University C              5             Data Science
3         4         101         4.0   2025-04-22  Blockchain for Security      University A              10          Blockchain
4         5         104         4.5   2025-04-21  Robotics in Manufacturing      University D              12          Robotics
5         6         102         3.0   2025-04-20      AI in Healthcare      University B              15             AI

13. Count of papers in each category:
Paper_Category
AI          2
Data Science  2
Blockchain  2
Computer Science  1
Robotics     1
Machine Learning  1
Cloud Computing  1

```

```

93 # 14. Find the average review score for papers from a particular university (e.g., University A).
94 avg_score_university_A = df[df['Reviewer_Affiliation'] == 'University A']['Review_Score'].mean()
95 print("\n14. Average review score for papers from University A:")
96 print(avg_score_university_A)
97
98 # 15. Find the reviewer with the most reviews.
99 most_reviews_reviewer = df.groupby('Reviewer_ID')['Paper_ID'].count().idxmax()
100 print("\n15. Reviewer with the most reviews:")
101 print(most_reviews_reviewer)
102
103 # 16. Find the paper with the highest review score for each reviewer.
104 highest_score_per_reviewer_paper = df.loc[df.groupby('Reviewer_ID')['Review_Score'].idxmax()]
105 print("\n16. Paper with the highest review score for each reviewer:")
106 print(highest_score_per_reviewer_paper)
107
108 # 17. Filter out papers that have a review score below 4.
109 papers_above_score_4 = df[df['Review_Score'] >= 4]
110 print("\n17. Papers with a review score above or equal to 4:")
111 print(papers_above_score_4)
112
113 # 18. Calculate the median review score for each paper.
114 median_score_per_paper = df.groupby('Paper_ID')['Review_Score'].median()
115 print("\n18. Median review score for each paper:")
116 print(median_score_per_paper)
117
118 # 19. Find the most recent paper reviewed (based on review date).
119 most_recent_paper = df.loc[df['Review_Date'].idxmax()]
120 print("\n19. Most recent paper reviewed:")
121 print(most_recent_paper)

```

# Output

```

14. Average review score for papers from University A:
4.0

15. Reviewer with the most reviews:
101

16. Paper with the highest review score for each reviewer:

```

Paper_ID	Reviewer_ID	Review_Score	Review_Date	Paper_Title	Reviewer_Affiliation	Reviewer_Experience (years)	Paper_Category	
0	1	101	4.2	2025-04-25	AI in Healthcare	University A	10	AI
1	2	102	3.5	2025-04-24	Quantum Computing Applications	University B	15	Computer Science
2	3	103	4.8	2025-04-23	Data Science in Business	University C	5	Data Science
4	5	104	4.5	2025-04-21	Robotics in Manufacturing	University D	12	Robotics
6	7	105	4.7	2025-04-19	Machine Learning for Cybersecurity	University E	8	Machine Learning
7	8	106	4.3	2025-04-18	Cloud Computing in Education	University F	6	Cloud Computing
9	10	107	4.6	2025-04-16	Blockchain for Security	University G	18	Blockchain

```

17. Papers with a review score above or equal to 4:

```

Paper_ID	Reviewer_ID	Review_Score	Review_Date	Paper_Title	Reviewer_Affiliation	Reviewer_Experience (years)	Paper_Category	
0	1	101	4.2	2025-04-25	AI in Healthcare	University A	10	AI
2	3	103	4.8	2025-04-23	Data Science in Business	University C	5	Data Science
3	4	101	4.0	2025-04-22	Blockchain for Security	University A	10	Blockchain
4	5	104	4.5	2025-04-21	Robotics in Manufacturing	University D	12	Robotics
6	7	105	4.7	2025-04-19	Machine Learning for Cybersecurity	University E	8	Machine Learning
7	8	106	4.3	2025-04-18	Cloud Computing in Education	University F	6	Cloud Computing
9	10	107	4.6	2025-04-16	Blockchain for Security	University G	18	Blockchain

```

18. Median review score for each paper:
Paper_ID
1    4.2
2    3.5
3    4.8
4    4.0
5    4.5
6    3.0
7    4.7
8    4.3
9    3.8
10   4.6
Name: Review_Score, dtype: float64

19. Most recent paper reviewed:
Paper_ID          1
Reviewer_ID      101
Review_Score      4.2
Review_Date      2025-04-25 00:00:00
Paper_Title      AI in Healthcare
Reviewer_Affiliation  University A
Reviewer_Experience (years)  10
Paper_Category    AI
Name: 0, dtype: object

```

```

123 # 20. Find the difference in review scores between two reviewers (e.g., Reviewer 101 and Reviewer 102) for the same paper.
124 df_pivot = df.pivot(index='Paper_ID', columns='Reviewer_ID', values='Review_Score')
125 score_diff_101_102 = df_pivot[101] - df_pivot[102]
126 print("\n20. Difference in review scores between Reviewer 101 and Reviewer 102 for each paper:")
127 print(score_diff_101_102)
128
129

```

## Output

```

20. Difference in review scores between Reviewer 101 and Reviewer 102 for each paper:
Paper_ID
1      NaN
2      NaN
3      NaN
4      NaN
5      NaN
6      NaN
7      NaN
5      NaN
6      NaN
7      NaN
5      NaN
6      NaN
5      NaN
5      NaN
5      NaN
6      NaN
7      NaN
8      NaN
9      NaN
10     NaN
dtype: float64
PS C:\Users\Dell\Desktop\EDS>

```